Genome-wide signatures of population bottlenecks and diversifying selection in European wolves

M Pilot[1,2], C Greco[3], BM vonHoldt[4], B Jędrzejewska[5], E Randi[3, 6], W Jędrzejewski[5, 9], VE Sidorovich[7], EA Ostrander[8], RK Wayne[4]

[1]School of Life Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK;

[2]Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, 00-679 Warsaw, Poland;

[3]Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), 40064, Ozzano Emilia (BO), Italy;

[4]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095, USA;

[5]Mammal Research Institute, Polish Academy of Sciences, 17-230 Białowieża, Poland;

[6]Aalborg University, Department 18, Section of Environmental Engineering, Aalborg, Denmark

[7]Institute of Zoology, National Academy of Sciences of Belarus, Minsk, Belarus;

[8]Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA;

[9]Present address: Instituto Venezolano de Investigaciones Cientificas (IVIC), Centro de Ecologia, Caracas, Venezuela

25 Correspondence: RK Wayne, Department of Ecology and Evolutionary
26 Biology, University of California, Los Angeles, California 90095, USA. E-
27 mail: rwayne@ucla.edu
28
29 Running title: Genome-wide diversification in European wolves

Abstract

Genomic resources developed for domesticated species provide powerful tools for studying the evolutionary history of their wild relatives. Here we use 61K single nucleotide polymorphisms (SNPs) evenly spaced throughout the canine nuclear genome to analyse evolutionary relationships among three largest European populations of grey wolves in comparison with other populations worldwide, and investigate genome-wide effects of demographic bottlenecks and signatures of selection. European wolves have a discontinuous range, with large and connected populations in Eastern Europe and relatively smaller, isolated populations in Italy and the Iberian Peninsula. Our results suggest a continuous decline in wolf numbers in Europe since the Late Pleistocene, and long-term isolation and bottlenecks in the Italian and Iberian populations following their divergence from the Eastern European population. The Italian and Iberian populations have low genetic variability and high linkage disequilibrium, but relatively few autozygous segments across the genome. This last characteristic clearly distinguishes them from populations that underwent recent drastic demographic declines or founder events, and implies long-term bottlenecks in these two populations. Although genetic drift due to spatial isolation and bottlenecks seems to be a major evolutionary force diversifying the European populations, we detected 35 loci that are putatively under diversifying selection. Two of these loci flank the canine platelet-derived growth factor gene, which affects bone growth and may influence differences in body size between wolf populations. This study demonstrates

the power of population genomics for identifying genetic signals of demographic bottlenecks and detecting signatures of directional selection in bottlenecked populations, despite their low background variability.

Keywords: bottleneck; effective population size; linkage disequilibrium; genetic differentiation; selection; grey wolf

INTRODUCTION

Studies on evolutionary processes in natural populations have been greatly enabled by technological advances related to whole genome sequence data from a variety of domesticated species (Allendorf et al. 2010). Access to large number of loci, often with annotated positions within the genome of the investigated species, permits researchers to overcome analytical limitations associated with the analysis of a small number of genetic markers. Examples include reconstruction of admixture patterns among closely related species (vonHoldt et al. 2011, Miller et al. 2012), identification of the genetic basis of parallel adaptations (Hohenlohe et al. 2010, Zulliger et al. 2013), and investigation of demographic effects of past climate change (Miller et al. 2012, Zhao et al. 2013). Here we use a population genomic approach to study the genetic effects of demographic bottlenecks in European grey wolf populations.

Demographic bottlenecks have been extensively explored using classical population genetic methods, typically based on a small number of neutral microsatellite loci (as reviewed in Peery et al. 2012), or MHC loci, presumably under balancing selection (e.g. Oliver & Piertney 2012). Given the limitations of using limited numbers of genetic markers (Peery et al. 2012), genome-wide studies based on data from natural populations that underwent population declines are needed. Considerable attention has been paid to population bottlenecks associated with domestication events and resulting problems with distinguishing true signals of selection from effects of drift (e.g. Caicedo et al. 2007, Axelsson et al. 2013). However, in

84 domestic species, a strong signal of artificial selection can be expected and
85 predictions can be made regarding traits likely to be affected, while in wild
86 species, the strength of selection and traits affected are less predictable.
87 Here we assess genome-wide effects of population bottlenecks and
88 identify signals of selection in European grey wolves (*Canis lupus*). The
89 grey wolf is the direct ancestor of the domestic dog (*Canis lupus familiaris*),
90 which is an important and emerging model for understanding the genetics of
91 disease susceptibility and developmental biology. Therefore, genomic
92 studies on the grey wolf benefit from the extensive genomic resources
93 available for the domestic dog (e.g. Lindblad-Toh et al. 2005, vonHoldt et
94 al. 2010, 2011). Another advantage of focusing on the grey wolf is the
95 extensive background knowledge regarding its ecology, recent demographic
96 history and population genetics (reviewed in Musiani et al. 2010, Randi
97 2011).
98 Genetic studies revealed a complex evolutionary history of the grey
99 wolf, with no clear phylogeographic patterns worldwide (Vilà et al. 1999,
100 Pilot et al. 2010), but with cryptic population genetic subdivisions related to
101 environmental differences (e.g. Geffen et al. 2004, Pilot et al. 2006,
102 vonHoldt et al. 2011). Wolves had a continuous range in Europe throughout
103 most of the Holocene, which was considerably reduced and fragmented in
104 the last few centuries as a result of direct eradication and habitat loss.
105 Currently, wolves in Western Europe occur in isolated and partially
106 protected populations in Italy (including the Apennine Peninsula and the
107 western Italian Alps) and the Iberian Peninsula. In Eastern Europe, there are

large and interconnected populations (Figure 1), most of which have experienced constant hunting pressures. Cryptic population structure has been observed in Eastern Europe (Pilot et al. 2006, Stronen et al. 2013), but this genetic differentiation is small compared to the differentiation between Eastern Europe and both Italian and Iberian populations. Therefore, herein, we use the term "Eastern European population" despite the lack of panmixia.

Patterns of mtDNA variability suggest that Eastern Europe and the Iberian Peninsula were linked by gene flow before the extinction of intermediate populations, a conclusion supported by the presence of a shared haplotype between the Eastern European and the Iberian population (Pilot et al. 2010). By comparison, long-term isolation has been suggested for the Italian wolf population (Lucchini *et al*. 2004) which has a unique mtDNA haplotype not found elsewhere.

The three main European wolf populations have distinct demographic histories. The Iberian Peninsula currently contains the largest wolf population in Western Europe, numbering over 2 000 individuals (Sastre et al. 2011). This population has been isolated at least since the extinction of the wolf from France at the end of the nineteenth century, and suffered a recent demographic bottleneck in the 1970s, when the population was reduced to about 700 individuals (Sastre et al. 2011). Since that time, the population has expanded in range and size. The current population has a small effective population size (about 50) and shows signs of the past genetic bottleneck (Sastre et al. 2011).

132 The Italian wolf population also experienced a severe demographic
133 bottleneck in 1970s, when it was reduced to about 100 individuals, the
134 effects of which are detectable at the genetic level (Randi 2011). However,
135 the history of this population may be more complex than a single recent
136 bottleneck. Lucchini et al. (2004) used a Bayesian coalescent analysis to
137 show that Italian wolves underwent a 100 to 1000-fold population
138 contraction during the last 2 000-10 000 years, which may be more
139 important in defining their current genetic profiles. As a result of recent
140 legal protection and abundance of prey, the Italian wolf has recovered to a
141 range that includes the entire Apennines and the western Italian Alps, and is
142 expanding to the Swiss and French Alps (Randi 2011), eastern Italian Alps
143 (Fabbri et al., in press) and even into Spain (Sastre 2011).
144 The wolf distribution in Eastern Europe is relatively continuous, and
145 is connected with Asian populations (Boitani 2003; Figure 1). To the best of
146 our knowledge, there is no account of any strong bottleneck that would
147 affect this population, although there is some evidence for a large-scale
148 population decline in the former Soviet Union and the neighbouring
149 European countries in the 1970's (Boitani 2003, Sastre et al. 2011).
150 However, the Eastern European population has experienced strong hunting
151 pressure for many generations, and the hunting continues in most of its
152 range to this day. As a result of hunting pressures on both the wolves and
153 their prey, the Eastern European wolves have suffered multiple local
154 demographic fluctuations (e.g. Spiridinov & Spassov 1985, Jędrzejewska et
155 al. 1996, Ozolins & Andersone 2001, Sidorovich et al. 2003, Gomercic et al.

156 2010).

157 Most of genetic studies on European grey wolves are based on a small
158 number of markers (nuclear and mitochondrial), with few comparative
159 studies across all the three populations (reviewed in Randi 2011). The
160 availability of validated tools for genome-wide analysis of SNPs in the
161 domestic dog opened new perspectives for population genetic studies of
162 wild canids (Lindblad-Toh et al. 2005). The utility of this approach has been
163 demonstrated by vonHoldt et al. (2011), who applied Affymetrix Canine
164 SNP Genome Mapping Array to study genome-wide variability in wild
165 wolf-like canids worldwide, with a focus on North America. That study
166 addressed long-standing questions about diversification and admixture in
167 wolf-like canids, including the systematic status of enigmatic taxa such as
168 the red wolf and Great Lakes wolf (vonHoldt et al. 2011). Here we analyse
169 genome-wide SNP variability in European grey wolves to test the following
170 hypotheses: (1) The three European populations should show high levels of
171 genetic differentiation, with the Italian population being particularly
172 distinct, reflecting its supposed ancient divergence and long-term isolation
173 (Lucchini et al. 2004, Pilot et al. 2010); (2) The Italian and Iberian
174 populations should show evidence for strong genetic bottlenecks (Lucchini
175 et al. 2004, Sastre et al. 2011); (3) A decline in effective size throughout the
176 last few centuries should be observed in each population as a result of a
177 direct extermination by humans and habitat loss (e.g. Randi 2011); and (4)
178 The three European populations should show a signal of diversifying
179 selection, reflecting their local adaptation to different types of habitat and

available prey (e.g. Geffen et al. 2004, Pilot et al. 2006, vonHoldt et al. 2011).

MATERIALS AND METHODS

Dataset

This study utilized data derived from the CanMap project (vonHoldt et al. 2010, Boyko et al. 2010) that provided genome-wide SNP data from 912 domestic dogs and 337 wild canids, based on genotyping with an Affymetrix Canine SNP Genome Mapping Array (coordinates based on the CanFam2 assembly). Samples were genotyped at 60 584 high-quality autosomal SNPs (referred to as 61K) and 851 X chromosome SNP loci (vonHoldt et al. 2010, Boyko et al. 2010). Here, we used a subset of the CanMap SNP dataset that consisted of 103 grey wolves: 54 from Eastern Europe, 19 from Italy, six from the Iberian Peninsula, seven from Asia, and 17 from North America, plus five coyotes that served as an outgroup.

For linkage disequilibrium and autozygosity analyses (see below), we introduced subdivision by defining small groups of spatially proximate samples within Eastern Europe (Figure 1B). These groups were delimited based on both geographical proximity of sampling locations and results of an earlier study showing genetic structure within Eastern Europe (Pilot et al. 2006), and therefore in some cases geographically proximate samples are assigned to different groups to reflect population differentiation found previously.

The initial set of 61K loci was pruned using PLINK (Purcell et al. 2007) for loci that were invariant among the sample set, or had very low minor allele frequency (MAF) (<0.01), resulting in 53 793 SNPs. For many applications, using a dataset pruned for loci in strong linkage disequilibrium (LD) is advised (e.g. Alexander et al. 2009). Therefore, we further pruned the dataset for SNPs with an $r^2 < 0.5$ within 50 SNP sliding windows, shifted and recalculated every 10 SNPs. This dataset consisted of 33 958 SNPs (referred to as 34K dataset).

Screening the dataset for related individuals

We screened the initial larger dataset for the presence of close relatives by calculating pairwise identity-by-state (IBS) estimates in PLINK. This approach alone was insufficient to identify all close relatives in the highly isolated and bottlenecked wolf populations from Italy and the Iberian Peninsula, as all pairs of individuals had IBS values >0.8, which in an outbred population is the empirical threshold for close relatives (vonHoldt et al. 2011). Therefore, for the Italian and Iberian populations, we identified close relatives using maximum-likelihood approaches as implemented in CERVUS 3.0 (Marshall *et al*. 1998) and KINGROUP 2 (Konovalov *et al*. 2004). CERVUS was used for parentage analysis, and KINGROUP was used to identify individuals related at the full-siblings and half-siblings level.

For CERVUS analysis, we selected loci with no missing data and with allele frequencies between 0.45 and 0.55. There were 827 SNPs that met those conditions in the Italian population and 1442 in the Iberian population.

For KINGROUP analysis, we randomly selected 100 SNPs from this set (which was the maximum number of loci accepted).

Using KINGROUP for the Italian population (initial N=23), we identified one pair of full-siblings and three pairs of half-siblings. Only one individual from each pair was retained in the dataset. Among the Iberian wolves (initial N=10), KINGROUP identified two pairs and one trio of full-siblings. CERVUS identified two parent-offspring pairs and one parent-offspring trio, consistent with three out of four full-sibling groups identified by KINGROUP, and only one individual from each pair or trio was retained in the dataset. The sample sizes after removing the closely related individuals were 19 for Italy, six for the Iberian Peninsula and 54 for Eastern Europe – this dataset was used in all the subsequent analyses.

Population structure analysis

1. Analysis of genetic differentiation in European wolves

We analysed the population genetic structure for the entire dataset consisting of European, Asian and North American grey wolves, with coyotes as an outgroup. Genetic structure analyses were performed using the 34K dataset. We used the Bayesian inference of genetic structure with no prior population information as implemented in STRUCTURE (Pritchard et al. 2000) and ADMIXTURE (Alexander et al. 2009). We used the two programs to check for consistency of the inferred structure.

STRUCTURE was run for K (the number of groups) from 1 to 10, with 100,000 MCMC iterations preceded by 20,000 burn-in iterations, and with

252 three replicates for each K value. We used the admixture model and
253 correlated allele frequencies. For each K, we checked whether the run
254 parameters (likelihood, posterior probability of data and alpha) reach
255 convergence within the burn-in period. Selection of optimal K based on
256 STRUCTURE output was performed with the support of STRUCTURE
257 HARVESTER software (Earl & vonHoldt 2012). We chose the optimal K
258 value based on likelihood values, the Evanno et al. (2005) ΔK method and
259 maximum biological information.
260 ADMIXTURE analysis was run for K from 2 to 10, using the default
261 termination criterion, which stops iterations when the log-likelihood
262 increases by less than $\varepsilon = 10^{-4}$ between iterations. The value of K for which
263 the model was optimally predictive was identified using a cross-validation
264 method in which runs are performed holding out 10% of the genotypes at
265 random, with 10 repetitions. The optimal K was selected as the value that
266 exhibited the lowest cross-validation error compared to other K values. We
267 also used ADMIXTURE to carry out a separate analysis for Eastern European
268 wolves only. We performed this additional analysis because earlier studies
269 suggested population structuring in this region (Pilot et al. 2006, Stronen et
270 al. 2013), which could have remained undetected in the context of strongly
271 differentiated wolf populations from other parts of the world.
272 Additionally, we performed a principal components analysis (PCA)
273 using the package SMARTPCA from EIGENSOFT (Patterson et al. 2006) to
274 visualize the dominant components of variability within the dataset. This
275 analysis was performed for: (1) the entire sample set; (2) European wolves;

and (3) Eastern European wolves. EIGENSOFT was also used to assess pair-wise F_{ST} and average divergence between and within populations (for details, see Supplementary Material).

2. Analysis of genetic structure using X chromosome data

The X chromosome data, which included 851 SNPs, were analysed for 37 females from the three European populations. We excluded SNPs from the pseudoautosomal region (PAR; first 6 Mb of the X chromosome). Outside the PAR, we removed additional four loci that were heterozygous in six males (which suggested genotyping errors). At each of the remaining 508 SNPs, no more than two males genotyped were heterozygotes. These were most likely genotyping errors and we treated them as missing data. After this adjustment, we obtained X chromosome haplotypes for males, which were used as a reference to improve the phasing of the corresponding female genotypes, which was carried out using FASTPHASE (Scheet and Stephens 2006). The inferred female haplotypes were used to construct a neighbour-joining tree in MEGA 5 (Tamura et al. 2011), using genetic distances calculated as the proportion of the number of different bases to the total number of SNP sites. This procedure was also carried out for 50 pure-breed domestic dogs available from the CanMap project (vonHoldt et al. 2010). We then selected 3 females to be included as an outgroup in the neighbour-joining tree. We also analysed population structure using ADMIXTURE (with the same parameter settings as described for the

299 autosomal data) for the LD-pruned X chromosome dataset consisting of 249
300 SNPs.
301
302 **Heterozygosity, linkage disequilibrium and autozygosity analysis**
303 We calculated observed and expected heterozygosity for the Iberian
304 and Italian wolf populations, and local populations from Eastern Europe
305 (see Figure 1B), based on the 61K SNP dataset. Because estimates of
306 heterozygosity and other parameters (see below) are dependent on sample
307 sizes, we included only the local populations with at least five individuals
308 sampled, and selected a random subset of six individuals from each of the
309 populations with more than six individuals. For these groups, we estimated
310 LD between all pairs of autosomal SNPs with MAF>0.15 by calculating
311 genome-wide pairwise genotypic association coefficient (r^2), based on the
312 61K SNP dataset. We estimated LD decay as the physical distance at which
313 r^2 coefficient decays below a threshold of 0.5.
314 Additionally, we identified runs of homozygosity (ROHs) >100 kb
315 spanning at least 25 SNPs in individuals from each population. Long ROHs
316 (>1 Mb) are indicative of autozygosity (i.e. homozygosity by descent) and
317 are a product of recent demographic events such as inbreeding or admixture,
318 whereas ROHs across shorter chromosome fragments (<1 Mb) are
319 indicative of more ancient population processes (Boyko et al. 2010).
320 Because our goal was to find ROHs that represent autozygosity rather than
321 simply occur by chance, this analysis was performed using the SNPs pruned
322 for local LD ($r^2 < 0.5$). In this case, the pruning was performed for each

local population separately. All the above analyses were performed in PLINK (Purcell et al. 2007).

Estimation of past demographic changes in European wolf populations

Effective population sizes (N_E) were estimated from the equation $E(r^2) = 1/(1+4N_E c) + 1/n$, where r^2 is a squared correlation in genotype frequencies between autosomal SNPs (representing the extent of LD), c is the genetic distance between loci in Morgans, and 1/n is the adjustment for small sample size (Tenesa et al. 2007). We assumed that 100Mb = 1 Morgan (as e.g. in Kijas et al. 2012). We estimated average values of r^2 in 20 distance classes between 2.5 kb and 1 Mb (corresponding to 0.0025 – 1 cM). We used the same distance classes as in the LD decay analysis (see Figure 5A), but the smallest distance class was not used here because r^2 estimates at small distances may be highly biased (Frisse et al. 2001, Gattepaille et al. 2013). Average r^2 value for a particular genetic distance (c) provides a N_E estimate t generations ago, where $t \approx 1/(2c)$ (Hayes et al. 2003). Therefore, the distance classes considered here translate into demographic changes from 50 to 20 000 generations ago, which corresponds to 150 – 60 000 years ago, assuming a generation time of 3 years (Mech & Seal 1987). The linear dependence between the recombination distance and time is approximate and holds best when population size is changing linearly (Hayes et al. 2003), which is not the case here (see Results). Therefore, the timing of the demographic changes being inferred here is approximate.

346 Temporal N_E changes were reconstructed for Eastern European
347 wolves (pooled), Iberian and Italian wolves. Because the correction for the
348 small sample size was applied, we did not use equal sample sizes, but
349 included all available individuals. However, we compared the results for the
350 Italian population based on 19 and 6 individuals and found them to be
351 similar (see Results). We also estimated the demographic changes for local
352 populations in Eastern Europe (as in the LD decay analysis) to compare
353 them with the global estimate for the entire Eastern European population. In
354 addition, the N_E estimates were also obtained for the North American
355 wolves. We expected them to have lower N_E estimates than Eastern
356 European wolves over time, because of a bottleneck (or, precisely, founder
357 effect) during the colonization of North America from Eurasia (Nowak
358 2003).
359
360 **Estimation of divergence times between the European wolf populations**
361 We used a method of Gautier & Vitalis (2013) implemented in the program
362 KIM TREE, which estimates divergence times on a diffusion time scale (i.e.
363 forward in time), conditionally on a population history that is represented as
364 a tree. The most likely tree topology is identified using the deviance
365 information criterion (DIC) (Spiegelhalter et al. 2002). The branch lengths
366 are estimated as $\tau \approx T/(2N_E)$, where τ is the length of the branch leading to a
367 particular population, N_E is the effective size of this population, and T is
368 time (in generations). We used this program to establish the order of
369 splitting events between the three European populations, and the relative

temporal distances between them. We also made an attempt to estimate divergence times in generation units as $2N_E\,\tau$, and then in years assuming a 3-year generation time. However, there was a considerable uncertainty connected with these estimates (see Supplementary Material).

Identification of candidate loci under selection

We used the program BAYESCAN (Foll & Gaggiotti 2008) to identify candidate loci under natural selection in European wolves. This analysis was performed for the three European populations (Eastern European, Iberian and Italian) using the entire 61K SNP set, but excluding loci that were monomorphic in European wolves (which gave 55 023 SNPs). BAYESCAN applies a Bayesian model developed by Beaumont & Balding (2004). It assumes an island model, where the difference in allele frequencies at each locus between each population and a common gene pool for all the populations is presented as a population-specific F_{ST}. Selection is introduced by decomposing F_{ST} coefficients for each locus into a population-specific component (ß) shared by all loci, and a locus-specific component (α) shared by all populations considered (Foll & Gaggiotti 2008). For example, three populations with a moderate level of genome-wide differentiation (e.g. average $F_{ST} = 0.1$), but fixed for three different alleles at a particular locus (locus-specific $F_{ST} = 1$) would have a high, positive value of α coefficient for this particular locus. Departure from neutrality is assumed for these loci for which the α–component is necessary to explain the observed pattern of diversity at a given locus. This

corresponds to α being significantly different from 0, with positive values suggesting diversifying selection, and negative values - balancing or purifying selection (Foll & Gaggiotti 2008). A threshold value to detect selection was set using a maximum False Discovery Rate (FDR; the expected proportion of false positives) at 0.05. This approach has been assessed as conservative in comparison with other methods of detecting selection (e.g. Zhao et al. 2013), but because of the nature of our data (bottlenecked populations) we did not use the relaxed FDR threshold of 0.1 applied elsewhere (e.g. Zhao et al. 2013). BAYESCAN accounts for the uncertainty of allele frequency estimates associated with small sample sizes, and therefore it can be applied for very small samples without bias, but with the risk of low power (Foll & Gaggiotti 2008). Therefore, our analysis has a low risk of detecting false positives, but it is likely that a number loci being under selection will remain undetected. For the SNPs identified as the candidate loci, we performed a search in UCSC Genome Browser for the closest protein-coding genes in CanFam2 dog genome assembly (SNP coordinates were based on this assembly), and also searched for homologous genes identified in humans and other mammals using this browser. Population differentiation at loci putatively under selection was assessed using the PCA implemented in the EIGENSOFT software.

RESULTS

Genetic differentiation among European wolf populations in relation to other Holarctic populations

Population genetic structure at genome-wide loci set

Both STRUCTURE and ADMIXTURE identified Italian wolves as the most distinct population at K=2, with North American canids (grey wolves and coyotes) identified as the third distinct group at K=3 (Supplementary Figure S1). The coyotes were not separated from wolves at K=2 because of the large differences in the sample sizes for these two groups (see Discussion). For larger values of K, the subsequent groups emerged in different order depending on the program used. ADMIXTURE identified coyotes as a distinct group at K=4, and STRUCTURE at K=7. ADMIXTURE identified K=6 as the most informative genetic subdivision, with the clusters corresponding to phylogenetic and geographic subdivision of the samples: Italian, Iberian, Eastern European, Asian and North American wolves, and coyotes. STRUCTURE identified K=7 as the most informative genetic subdivision, both based on the maximum likelihood and the Evanno et al. (2005) method. The clusters identified were the same as in ADMIXTURE for K=6, but with one additional cluster that was represented in most Eastern European individuals as a secondary genetic component.

This cluster constituted the main component of the genetic variability for only three individuals from the Carpathian Mountains, with eight other individuals from the Carpathian Mountains and the Balkans showing levels of admixture with this cluster: between 0.27 and 0.48. The same cluster was identified in ADMIXTURE at K=7. The differences in assignment

probabilities to these clusters may suggest further differentiation between the Carpathian Mountains and the Balkans. Some Eastern European individuals, in particular those from the easternmost sampling area, i.e. the Kirov Region in Russia (see Figure 1B) showed mixed ancestry with Asian wolves (Figure 2). Although these results suggest some level of differentiation within Eastern Europe, the separate analysis including only Eastern European wolves detected no population structure (see Supplementary Material), which may be a result of uneven sample distribution and small sample sizes (see Discussion).

Principal Component Analysis

In the analysis including grey wolves from Europe and other continents as well as coyotes, the first axis (PC1; 8.8% variation) discriminated Italian wolves from other populations (Figure 3A). From positive to negative values, the second axis (PC2, 5.7% variation) separated coyotes, North American grey wolves, Italian wolves, Asian wolves and other European wolves. A similar trend of decreasing values on PC2 was also observed within Eastern European wolves, with individuals from regions geographically more proximate to Asia (from easternmost sampling locations in Russia and Ukraine) placed closer to Asian wolves in the PCA plot. Iberian wolves clustered with Eastern European wolves, but with some separation. The level of differentiation between Eastern European and Italian wolves (F_{ST}=0.195) was higher than that between Eastern European and North American wolves (F_{ST}=0.114; Figure 3A).

The analysis including only European wolves revealed that PC1 (8.3% variation) separated Italian wolves from Eastern European and Iberian wolves, while PC2 (2.9% variation) separated Iberian wolves from the other populations. Within Eastern European wolves, individuals from the Balkans (Bulgaria, Croatia and Greece) and northeastern Europe (Belarus, Latvia, Poland, Russia, Slovakia and Ukraine) formed two distinct subclusters (Figure 3B).

In the analysis including only Eastern European wolves, PC1 (2.1% variation) separated wolves from the Carpathians, the Balkans, and northeastern Europe. PC2 (1.8% variation) separated different groups from northeastern Europe, but they were not geographically clustered.

Genetic differentiation among populations

As expected, the highest pair-wise F_{ST} values were observed between the coyotes and the grey wolves (Table 1A). Among wolves, the highest F_{ST} value (0.293) was observed between Italian and Iberian populations, whereas the lowest F_{ST} value (0.059) was observed between Eastern European and Asian populations. Eastern European population was more divergent from the Italian and Iberian populations than from the Asian and North American populations (Table 1A).

Average divergence values between populations did not follow the same pattern as F_{ST}, which was due to the lack of correction for intra-population divergence. Intra-population divergence was low in the Italian and Iberian populations (0.609 and 0.871, respectively), high in the Asian

population (1.623), and had intermediate values in Eastern European and North American populations (Table 1A). Genetic differences between each pair of populations were significant (ANOVA, $P<0.0001$).

Population structure based on X chromosome data

The neighbour-joining tree of female X chromosome haplotypes showed that Italian and Eastern European wolves were grouped in two distinct clades, but only the clade of Italian wolves was supported by the bootstrap analysis (Figure 4). Iberian haplotypes were clustered with Eastern European wolves, forming a distinct subclade with 100% support. There was no clear geographical structure among Eastern European wolves (Figure 4). In the majority of cases, the two X chromosome haplotypes of individual female wolves were not placed next to each other in the tree. The exceptions where two haplotypes of the same individuals were more similar to each other than to any other haplotype included two (100%) Iberian wolves, two (18%) Italian wolves, and two (8%) Eastern European wolves.

ADMIXTURE analysis for female X chromosome data distinguished Italian wolves from Eastern European wolves at K=2, with the two Iberian individuals grouped with Eastern European wolves. At K=3 (which was indicated as the most likely genetic structure), two clusters were identified for Eastern European wolves: one comprised of Carpathian and Balkan individuals, and second of the individuals from northeastern Europe. These clusters were consistent with those detected using the autosomal loci set.

Heterozygosity, linkage disequilibrium and autozygosity in local populations of European wolves

All local wolf populations from Eastern Europe had comparable levels of heterozygosity (H_O = 0.21 - 0.24, H_E = 0.22 - 0.26; Table 2), whereas populations from southwestern Europe exhibited lower heterozygosity (Iberian Peninsula: H_O and H_E = 0.17; Italy: H_O and H_E = 0.16). Eastern European wolves had low to moderate levels of LD (LD decayed below r^2=0.5 between 2.5 and 10 Kb), as expected for populations that have not experienced severe bottlenecks. The southernmost population from the Balkans had the highest LD levels within Eastern European populations (Figure 5A). In contrast to these populations, the Iberian population had high LD levels (257 Kb), consistent with bottlenecks and subsequent inbreeding. In the Italian population, LD did not decay below 0.5 for the entire range of distances considered (up to 1 Mb), suggesting more severe and/or a longer bottleneck as compared to the Iberian population (Table 2, Figure 5).

Despite high LD levels in the Iberian and Italian populations, they had fewer fragments of ROH > 1 Mb than Eastern European populations (Table 2, Figure 5). In contrast, some Eastern European populations, such as the Carpathians, Northern Belarus, and Southern Russia/Eastern Ukraine, had an elevated number of ROH fragments of smaller size (1-5 Mb).

Past demographic changes in European wolf populations

LD-based estimates suggest that effective population sizes of both European and North American wolves declined over the entire period considered (Figures 5C and S2). N_E estimates for the Italian and Iberian populations were considerably lower as compared with the Eastern European population in each time interval (Figures 5C and S2). The most recent effective population sizes (at about 150 years ago) were estimated at 1366 for Eastern Europe, 71 for Italy, and 59 for the Iberian Peninsula. The most ancient estimates (at about 60 000 years ago) were: ~20 000 for Eastern Europe, ~4500 for Italy and ~10 000 for the Iberian Peninsula (Table S2). Prior to the divergence of the European populations (which most likely occurred within the timeframe considered), their N_E should be the same, which is not observed. This may be interpreted as an evidence for long-term bottlenecks in the Italian and Iberian populations (see Discussion) or ancient population structure. N_E estimates for the North American wolves (most recent: 358, most ancient: ~18 000) do not converge on those of Eastern European wolves, either, which may reflect a more complex demographic history of North America, including multiple founder effects and bottlenecks associated with glaciation events (Nowak 2003). Most local groups of Eastern European wolves do not converge to the effective size of the total Eastern European population (Figure 5C), which may result from population structure (Pilot et al. 2006) and/or local bottlenecks.

Divergence times between the European wolf populations

559 The most likely tree topology inferred using the Kim Tree program suggests
560 that the Iberian population diverged first from the common ancestor of all
561 populations considered, which was followed by the split between the Italian
562 and Eastern European populations (Figure S3A). However, small
563 differences in DIC values (92-227) between the alternative topologies and a
564 very short internal branch (Figure S3A) suggest that the splits between these
565 three populations occurred within a short time period, and the topology is
566 close to star-shaped.
567 We added North American grey wolves to the most likely tree
568 topology of the three European populations, assuming the reciprocal
569 monophyly between European and North American wolves (as shown in
570 vonHoldt et al. 2011). Using the time of flooding of the Bering Land Bridge
571 (11 000 yBP, Keigwin et al. 2006), which separated Eurasian and North
572 American wolves as a calibration point, we obtained the conservative
573 estimates of divergence of European populations from their most recent
574 common ancestor at 3200-5600 years ago (SD 33-123 years) (Table S3,
575 Figure S3B). These estimates have considerable uncertainty resulting from a
576 number of assumptions (see Supplementary Material for details), and
577 therefore should be treated with caution.
578
579 **Identification of candidate loci under selection**
580 Using a 5% FDR threshold, we identified 35 outlier SNPs (Figure 6,
581 Table S4). This threshold corresponded to Posterior Odds (PO) of 8.94 and
582 False Non-Discovery Rate (the expected proportion of false negatives) of
583 0.094, and P=0.90, respectively. Thirty-one of these outliers fitted within a

584 threshold of PO<10. Each of the 35 outliers had positive α values between
585 1.27 and 2.03, suggestive of diversifying selection. F_{ST} coefficient averaged
586 over populations ranged from 0.45 to 0.58 compared to the average value of
587 0.21 among the genome-wide 34K loci (Figure S4). None of these 35 outlier
588 loci showed evidence for directional selection within Eastern European
589 wolves (see Supplementary Material).

590 A search of the CanFam2 dog genome assembly in the UCSC
591 Genome Browser for the closest protein-coding genes indicated that two
592 outlier SNPs from chromosome 6 are flanking the coding region of platelet-
593 derived growth factor, alpha polypeptide (*PDGFA*). The first SNP (further
594 referred to as locus *PDGFA-1*) was 4.6 Kb downstream from the
595 chromosomal fragment marked as a coding region, and the second SNP
596 (*PDGFA-2*) was 30.7 Kb upstream. Locus *PDGFA-1* was among 5 loci with
597 PO > 100 (corresponding to $P > 0.99$) and had the highest α value of all loci
598 (α=2.03) and the highest level of differentiation (F_{ST}=0.58). Additionally,
599 one more putatively selected locus (*PDGFA-3*) was 425 Kb upstream from
600 this chromosomal fragment.

601 For the remaining 32 SNPs identified as putative loci under
602 selection, we found adjacent regions analogous to genes described in the
603 human genome and other mammalian genomes (Table S4), which have not
604 been annotated for the dog yet. One of these loci, which had the highest PO
605 value and second highest F_{ST} of all the loci putatively under selection, was
606 placed within a sequence analogous to thrombospondin type 1 gene
607 (*THBS1*), which was annotated in humans, mice and rats. Another locus was

placed within a sequence analogous to metallopeptidase with thrombospondin type 1 motif (*ADAMTS3)*, which was annotated in humans, mice, rats, and cows. Functions of thrombospondin type 1 include angiogenesis, apoptosis, and activation of transforming growth factor beta *(TGFβ)*.

Population differentiation at loci putatively under selection

The PCA plot representing genetic differentiation among worldwide grey wolf populations and coyotes at loci putatively under selection showed different pattern as compared with that obtained for the 34K dataset. While separation of Italian wolves from other wolves and coyotes at PC1 was consistent with the 34K dataset, at PC2 Iberian wolves were the most distinct population, and they were more similar to coyotes than to Eastern European wolves (Figure 3D). There was no clear distinction between Eastern European, Asian and North American wolves. At PC1, *PDGFA-3* and *PDGFA-1* were the first and the third of loci showing the highest level of differentiation among populations. On PC3, which distinguished the coyotes from the grey wolves, *PDGFA-2* showed the second highest level of differentiation among populations, after another locus from the same chromosome, but more distant from *PDGFA* gene.

The PCA plot representing the differentiation among the three European populations at loci putatively under selection using the PCA method showed a similar pattern as compared with the differentiation at 34K loci (Figure 3E), but as expected, with much stronger differentiation.

For example, PC1 (distinguishing Italian and Eastern European wolves) explained 42.9% of genetic variation versus 8.3%. Similarly, PC2 (distinguishing Iberian wolves from the two other populations) explained 5.7% of genetic variation versus 2.9%. Differentiation between northeastern and southeastern Europe observed for the 34K dataset was not observed here. At PC1, two loci showing the highest level of differentiation among populations were *PDGFA-1* and *PDGFA-3*. At PC3, which distinguished Italian wolves from other European wolves, *PDGFA-2* showed the highest level of differentiation among populations.

As expected in case of diversifying selection, pair-wise F_{ST} values between the European populations were highly elevated (0.758-0.925), with the highest level of differentiation between the Italian and Iberian populations. A less obvious effect was the substantial elevation of F_{ST} values between the coyotes and each of the grey wolf populations (0.359-0.867; Table 1B).

DISCUSSION

Genetic differentiation among European wolf populations in relation to other Holarctic populations

We detected six genetically distinct groups within the analysed dataset, which were consistent with species-level and geographic subdivision of the samples. Specifically, Italian, Iberian, Eastern European, Asian and North American wolves, and coyotes formed distinct clusters. The population structure based on X chromosome haplotypes confirmed the

high level of genetic differentiation among the three main European populations.

The genetic distinctiveness of the Italian, Iberian and Eastern European populations was expected given their geographic isolation and likely near complete lack of gene flow for at least last 100 years (Lucchini et al. 2004), except for the last decade of wolf population expansion in Western Europe (Sastre 2011, Fabbri et al., in press). These three populations spatially correspond to different glacial refugia: the Apennine and Iberian refugia for the two southwestern populations, and the Balkan refugium for the southeastern (Balkan) population (with northeastern European population possibly having a different or mixed origin - see Pilot et al. 2010). It has been unclear, though, whether the distinctiveness of these populations results from their long-term isolation or recent geographical separation resulting from extinction of the wolf in central-western Europe. The wolf range in Europe during the Last Glacial Maximum was not reduced to the southern refugia (see Sommer & Benecke 2005), so the effect of Pleistocene glaciations on population structuring in this species may be overestimated. Our estimates support the ancient divergence of the three European populations (5600-3200 years ago), but this date is considerably later than the Last Glacial Maximum (~20 000 years ago). Although this estimate has considerable uncertainty, when combined with other evidence it implies a new hypothesis concerning the events leading to the divergence of these populations (see below).

All the methods of population structure analysis indicate that the Italian population is the most genetically distinct of the three European populations considered here. This is consistent with an inference based on mtDNA data from modern and ancient European wolves, suggesting historic gene flow between Eastern Europe and the Iberian Peninsula through intermediate populations, and longer-term isolation of wolves in the Apennine Peninsula (Pilot et al. 2010). An analysis based on microsatellite loci also suggested the isolation of the Italian wolf population for thousands of generations (Lucchini et al. 2004).

In contrast, the genealogy of the European populations inferred using the Kim_Tree method suggests that the Iberian population diverged first from the ancestral European population, which was followed by the divergence between the Italian and Eastern European populations. However, the support for this tree topology over the alternative topologies is weak, and the internal branch is short, suggesting that the splits between these three populations occurred within a short period, and the topology is close to star-shaped.

The effect of sampling on the analysis of population structure

PCA suggested some level of differentiation within the Eastern European wolves, as the Carpathian, Balkan and northeastern populations formed distinct sub-clusters. However, this separation was not well supported by Bayesian clustering methods. These methods detected two genetic clusters within Eastern Europe, one prevailing in northeastern

703 Europe and another in the Carpathians and the Balkans, but with high level
704 of admixture. Lack of clear, geographically clustered genetic subdivision
705 within the Eastern European wolves contrasted with an earlier study that
706 showed cryptic population structure in this region based on 14 microsatellite
707 loci and mtDNA variability (Pilot et al. 2006), which was subsequently
708 confirmed based on an independent sample set collected from a smaller area
709 (Czarnomska et al. 2013). The discrepancy is likely due to much lower
710 sample coverage, as 54 Eastern European wolves were analysed here versus
711 643 wolves in Pilot et al. (2006). In this case, the result based on a small
712 number of loci, but large sample size is more reliable, which demonstrates
713 the importance of the sample size in population structure studies, regardless
714 of the number of loci.

715 The effect of sample size was also evident in the analysis of coyote
716 data. Although their distinctiveness from grey wolf populations was clearly
717 reflected in pair-wise F_{ST} values, it was less clear based on PCA and
718 population structure plots. Because grey wolves predominated in the sample
719 and only five coyotes were tested, hence subdivisions within grey wolves
720 dominated the results. By comparison, a study that analysed the same SNP
721 data with a more balanced numbers of grey wolves and coyotes (vonHoldt
722 et al. 2011) identified a clear distinction between these species, consistent
723 with past phylogenetic studies (e.g. Vilà et al. 1999; Lindblad-Toh et al.
724 2005). This is consistent with the simulation study showing that variation in
725 sample size may affect the population clustering inferred in STRUCTURE
726 (Kalinowski 2011).

Although small sample sizes may affect the reliability of genetic structure analysis, the availability of a large number of loci with uniform genome-wide distribution enables other analyses that are largely independent of the sample sizes. Genome-wide data proved to be very effective in reconstructing past demographic changes and detecting signatures of selection based on small sample sizes (e.g. Jones et al. 2012, Keller et al. 2013), which may be reduced even to single individuals when high-coverage genome sequences are available (e.g. Miller et al. 2012, Zhao et al. 2013, Freedman et al. 2014).

Genetic diversity and linkage disequilibrium in European wolf populations: detecting genome-wide signatures of population bottlenecks

Eastern European wolves had levels of heterozygosity comparable with large grey wolf populations from Canada and northwestern United States, which have a history of constant or recently expanding population size (vonHoldt et al. 2011). Italian and Iberian wolves had decreased heterozygosity and higher LD levels as compared with Eastern European wolves, which is consistent with earlier studies that reported signatures of bottlenecks in these populations based on microsatellite loci analysis (Lucchini et al. 2004, Sastre et al. 2011). Despite high LD levels, both the Italian and Iberian population had fewer ROHs over 1 Mb in length as compared to Eastern European wolves, suggesting that the high LD levels are likely due to ancient bottlenecks rather than recent inbreeding.

751 Consistent with this result, the levels of observed and expected
752 heterozygosity were comparable in both the Italian and Iberian population,
753 while in the recently bottlenecked Mexican wolf population observed
754 heterozygosity was much lower than expected (0.12 versus 0.18), implying
755 recent inbreeding (vonHoldt et al. 2011).
756 LD levels in Italian and Iberian populations were also lower as
757 compared to Mexican wolves and a small, isolated, recently founded
758 population from the Isle Royale National Park (vonHoldt et al. 2011). These
759 two North American wolf populations also had the highest fraction of
760 autozygous segments across all chromosomal fragment sizes of all
761 populations of North-American wolf-like canids (vonHoldt et al. 2011).
762 This contrasts with the Italian and Iberian wolves, for which autozygosity
763 levels are low compared with Eastern European populations. The analysis of
764 genome-wide variability thus shows a clear distinction between populations
765 that are inbred due to recent drastic demographic declines or founder events
766 such as the Mexican and Isle Royale wolves, respectively, as compared with
767 populations that have reduced levels of genetic variability due to long-term
768 isolation and low population sizes lasting for a large number of generations
769 such as the Italian and Iberian wolves. Populations with these two different
770 types of demographic history have been designated as “bottlenecked”. Here
771 we show that there is a clear difference in the genomic signature of their
772 demographic histories. This result has important implications for studies
773 where a genetic analysis is the only source of information on demographic
774 history.

775 Analysis of phylogenetic relationships among female X
776 chromosome haplotypes showed that all the female wolves from the Iberian
777 Peninsula and two of the 11 female wolves from Italy had haplotypes that
778 were more related to each other than to any other haplotype. This suggests
779 that these populations have an increased probability of forming mating pairs
780 between individuals sharing a recent common ancestry (even if not directly
781 related). This is expected for populations that have experienced isolation
782 and long-term bottlenecks. In contrast, in Eastern Europe, only two out of
783 24 individuals carried X chromosome haplotypes showing close
784 phylogenetic similarity, while in other cases haplotypes from distant
785 locations were phylogenetically related. This result is consistent with
786 substantial gene flow between different parts of Eastern Europe, which may
787 counterbalance the effects of recent local inbreeding (see below).
788 The Italian population had lower variability as compared with the
789 Iberian population (although more individuals were analysed), consistent
790 with earlier studies based on mtDNA and microsatellite loci (Vilà et al.
791 1999, Pilot et al. 2010, Sastre et al. 2011). Moreover, the Italian population
792 had higher LD levels as compared with the Iberian population, an indication
793 of longer and/or more severe bottleneck events in Italian wolves. This
794 finding is consistent with the conclusion based on population structure
795 analyses, and with an earlier study suggesting long-term isolation of the
796 Italian population based on microsatellite data (Lucchini et al. 2004). In
797 contrast, mtDNA haplotype sharing between Iberian and Eastern European
798 wolves suggested more recent gene flow between Iberian and Eastern

European wolves, most likely through now-extinct intermediary populations (Pilot et al. 2010). The present study showed high pair-wise population divergence estimates between Eastern European population and both Italian and Iberian populations, and the divergence between the Italian and Iberian populations is highest of all pairs of the wolf populations studied. This inconsistency between genetic and geographical distance may be a result of strong genetic drift during population bottlenecks in the Iberian and Apennine Peninsulas. In contrast with the Italian and Iberian populations, wolves from some Eastern European regions had elevated levels of ROH, suggesting recent inbreeding. This was likely connected with the disruption of pack structure due to strong hunting pressure (e.g. see Jędrzejewski et al. 2005). In one of the regions with elevated ROH levels, Northern Belarus, strong hunting pressure has been well documented (Sidorovich et al. 2003).

Past demographic changes in European wolf populations

Effective population sizes of European and North American wolves inferred from LD patterns decline over the entire period considered (60 000 to 150 years ago). This is consistent with the growing evidence from ancient DNA studies showing that large mammal species experienced a considerable loss of genetic diversity since the late Pleistocene (reviewed in Hofreiter & Barnes 2010). In particular, the loss of mtDNA haplotypes has been documented in North American (Leonard et al. 2007) and European grey wolves (Pilot et al. 2010), and this was correlated with the loss of

822 morphological and ecological diversity (Leonard et al. 2007, Germonpré et
823 al. 2009).
824 While a general trend of N_E decline in time is consistent with the
825 expectation, we also expected a signal of population growth after the Last
826 Glacial Maximum reflecting the spatial expansion to the areas previously
827 covered by the retreating ice sheet. The spatial expansion has been
828 documented based on the sub-fossil record (Sommer & Benecke 2005), but
829 it is possible that it was not accompanied by a substantial demographic
830 expansion, e.g. due to declines of large herbivore prey (see Hofreiter &
831 Barnes 2010) and exponential growth of the human population (see e.g.
832 McEvoy et al. 2011). The demographic reconstruction based on high-
833 coverage genome sequences shows a continuous decline of wolf populations
834 in Europe, Middle East and East Asia since ~20 000 years ago until present
835 (Freedman et al., 2014). This is consistent with our result, but also shows
836 that our upper time limit of 60 000 years for the decline may be
837 overestimated due to an imprecision of time estimates based on
838 recombination distance.
839 N_E estimates in the most recent time period considered (~150 years
840 ago) show a good correspondance with estimates for the contemporary (21st
841 century) populations. Sastre et al. (2011) reports N_E ~50 (43-54) for the
842 contemporary Iberian population, which corresponds well with our N_E
843 estimate of 59 individuals about 150 years ago. The contemporary N_E
844 estimate for northeastern part of European Russia (138-312.5; Sastre et al.
845 2011) is also consistent with our estimates for three local populations from

846 this region (159 in NE Russia, 224 in S Russia/E Ukraine and 239 in N
847 Belarus). Importantly, the contemporary N_E estimates result in N_E to census
848 size ratio of about 0.11 in Russia and 0.025 in the Iberian Peninsula,
849 suggesting a severe bottleneck and/or an overestimation of the current
850 census size in the Iberian population (Sastre et al. 2011).
851 Prior to the divergence of the European populations (which took
852 place within the considered timeframe – see below), their N_E estimates
853 should converge, which is not observed. In the analogous analysis carried
854 out for humans, N_E estimates for non-African populations are lower than
855 those of African populations instead of converging to the same values prior
856 to the divergence time (McEvoy et al. 2011). This pattern was interpreted as
857 a signature of the "out of Africa" bottleneck (McEvoy et al. 2011). A drastic
858 reduction of population size inflates r^2 estimates even for the small distance
859 classes (representing distant time periods), leading to an underestimation of
860 N_E before the bottleneck (McEvoy et al. 2011). Therefore, the patterns
861 observed in the Italian and Iberian populations may be interpreted as an
862 evidence for bottlenecks, with the more severe bottleneck in the Italian
863 population as compared with the Iberian population.
864 The timing of these bottlenecks cannot be inferred from the LD
865 patterns. Continuous population decline observed for each population
866 suggests that there was no recovery phase which would have marked the
867 end of the bottleneck period. However, the timing of a strong bottleneck is
868 expected to coincide with coalescence of lineages involved in this
869 bottleneck, resulting in a genealogy with short internal branches close to the

root (Gattepaille et al. 2013). The genealogy reconstructed for the European wolves has this topology, so it may be expected that the time of their divergence corresponds with a bottleneck period, or with an onset of a long-term bottleneck. This time was estimated at 5600-3200 years ago, which corresponds to the late Neolithic in Europe.

There is a considerable uncertainty associated with this estimate, resulting from a number of assumptions made. For example, we made an unrealistic assumption that there was no or little gene flow between the populations after the split, and therefore the divergence times are likely to be underestimated (see Gautier & Vitalis 2013). However, in consistence with other evidence from this and earlier studies (e.g. Lucchini et al. 2004), this estimate shows that the population bottlenecks in Italian and Iberian wolves were ancient rather than recent. Possibly, they could have resulted from the Neolithic expansion of the human population (e.g. Bocquet-Appel 2011) leading to increased hunting pressure and competition for resources (large game species) with humans, as well as habitat loss due to agricultural expansion. Human population growth and habitat loss have continued until present, preventing the recovery of wolf populations from past bottlenecks, which may explain the observed pattern of continuous decline. Contemporary expansion of the wolf populations in Europe (e.g. Boitani 2003, Randi 2011), largely resulting from their release from hunting pressure, is too recent to be detected from LD patterns.

Signatures of diversifying selection among European populations

894 In populations that have experienced recent bottlenecks, large numbers of
895 loci may display low levels of heterozygosity as a result of genetic drift, and
896 therefore directional selection may be difficult to detect (e.g. Axelsson et al.
897 2013). To account for this problem, we considered outliers in the empirical
898 distribution as candidate targets of selection, and established a conservative
899 outlier threshold. In addition, we compared variation at putatively selected
900 loci in the populations for which selection test has been performed to that in
901 non-tested populations, expecting that signatures of selection will be
902 consistent across multiple populations or across closely related species, as
903 has been shown in other studies (e.g. Hohenlohe et al. 2010, Zulliger et al.
904 2013).
905 We identified 35 putative loci under diversifying selection among 55K
906 SNPs tested. These estimates are conservative and associated with a nearly
907 10% false non-discovery rate. For most of the outlier SNPs, appropriately
908 annotated genome data was unavailable and as a result, associations with
909 particular genes are uncertain. However, three outlier SNPs were flanking
910 the coding region of the canine platelet-derived growth factor, alpha
911 polypeptide (*PDGFA*) gene. The presence of these three loci with the strong
912 signature of selection near this gene (one of which had the highest F_{ST} from
913 all the loci analysed; Figure 6) makes it a strong candidate gene under
914 diversifying selection among wolf populations. This gene takes part in
915 numerous developmental processes (Alvarez et al. 2006). Importantly, it
916 interacts with insulin-like growth factor-1 (*IGF1*) in the development of
917 bone and cartilage tissues, which was described in humans (e.g. Schmidt et

al. 2006, Bassem & Lars 2011) and dogs (Stefani et al. 2000). Sutter et al. (2007) found that a single allele of the *IGF1* gene determines small size in dogs and this gene shows a signature of intense artificial selection. The small size allele was absent from a large worldwide sample of grey wolves (Gray et al. 2010), and we found no signature of selection on *IGF1* in wolves. Consequently, rather than *IGF1*, *PDGFA* may be a major gene influencing body size differences observed in European grey wolves (see below). However, it should be noted that differences in body size between wolf populations across Europe are small as compared with differences between dog breeds.

Additionally, a SNP that had the highest PO value was placed within a sequence analogous to human thrombospondin type 1 gene, and another SNP was located within a sequence analogous to human *ADAMTS3* gene with thrombospondin type 1 motif. Thrombospondin type 1 takes part in a number of developmental processes, including activation of *TGFβ*, another growth factor produced by platelets and involved in bone development (e.g. Reddi & Cunningham 1990). Thus, diversifying selection on the European wolf populations appears to involve two different growth factors that possibly may be associated with differentiation of body size and shape.

The Italian and the Iberian wolf have been recognized as separate subspecies *Canis lupus italicus* (Altobello 1921) and *Canis lupus signatus* (Cabrera 1907) based on morphological differences including overall body size, coat coloration, and cranial measurements (Cabrera 1907, Altobello 1921, Vilà 1993, Nowak & Federoff 2002). Although body size differences

942 across Europe are not large (Vilà 1993) and may be due to phenotypic
943 plasticity or genetic drift resulting from long-term isolation, it is also
944 possible that they reflect local adaptation. Smaller body size in grey wolves
945 may have a selective advantage in habitats with smaller prey (MacNulty et
946 al. 2009), and the three European populations occupy distinct habitats that
947 differ in species composition and the relative abundance of ungulate prey.
948 Smaller species like the roe deer (*Capreolus capreolus*) and the wild boar
949 (*Sus scrofa*) are common in the wolf diet in the Iberian Peninsula and Italy
950 (e.g. Barja 2009, Mattioli et al. 2011), whereas larger prey such as the red
951 deer (*Cervus elaphus*) and moose (*Alces alces*) are more frequent in the
952 wolf diet of northeastern Europe (Jędrzejewski et al. 2010).
953 Importantly, although selection was inferred using the European
954 dataset only, the patterns of population differentiation at putatively selected
955 loci among worldwide grey wolves and coyotes were substantially different
956 when compared with that obtained for the 34K dataset. Particularly striking
957 is the position of the coyotes on the PCA plot (Figure 3D), showing reduced
958 relative distance between this species and Iberian and Italian wolves as
959 compared with the 34K dataset. Coyotes are smaller than North American
960 grey wolves, feed on smaller prey species and their natural geographic range
961 was south of the grey wolf range (Gompper 2002). Therefore, parallel
962 patterns of diversifying selection may exist among European grey wolves
963 and North American large canids. The contrasting pattern between the
964 putatively selected loci and genome-wide loci may reflect parallel
965 adaptation involving the same genes. Further study is required to assess the

966 role of these candidate genes in the adaptive diversification of wolf-like
967 canids, which could involve DNA and protein sequence characterization in
968 multiple populations, analysis of gene expression, quantitative analysis of
969 relevant phenotypic traits, and possibly functional in vitro studies.
970
971 **Conclusions**
972 Our analysis of genome-wide variability provided new insights into
973 the evolutionary history of the grey wolf in Europe, revealing continuous
974 population declines since the Late Pleistocene as well as long-term isolation
975 and demographic bottlenecks in southwestern Europe. Eastern European
976 wolves show more genetic similarity to Asian wolves than to Italian and
977 Iberian wolves, and Italian wolves are particularly distinct from other wolf
978 populations. This patterns results from strong genetic drift and does not
979 reflect phylogenetic relationships among lineages. The Italian and Iberian
980 populations show the genomic signature of long-term bottlenecks, which is
981 clearly different from recent drastic population declines or founder events
982 such as in the Mexican and Isle Royale wolves. The fact that these
983 demographic histories can be distinguished based on genomic data may be
984 important in cases where genetic variability is the only source of
985 information.
986 We detected 35 loci putatively under diversifying selection between
987 the three main European populations. Two of these loci were within 31 Kb
988 from the canine PDGF gene which may influence differences in body size
989 between wolves from eastern and southwestern Europe. The contrasting

pattern of genetic differentiation among the populations of grey wolves and the coyotes at the putatively selected versus genome-wide loci may reflect parallel adaptation involving the same genes, a possibility that should be explored by resequencing studies of both species.

DATA ARCHIVING

The genotyping data from the CanMap project are available at http://genome-mirror.bscb.cornell.edu/cgi-bin/hgGateway (see ‘‘SNPs’’ track under the Variations and Repeats heading).

ACKNOWLEDGMENTS

We thank M. Shkvyrya, I. Dikiy, E. Tsingarska, S. Nowak and M. Apollonio for providing samples from European wolves used in this project, A. Moura for help with preparing the figures, and M. Gautier for sharing the R script to plot the outputs from the Kim_Tree program. We are grateful to Giorgio Bertorelle, Michael Bruford and four anonymous reviewers for their constructive comments on the manuscript. This project was supported by grants from the Foundation for Polish Science (M.P.), the Polish Committee for Scientific Research (M.P. and W.J.), the US National Science Foundation (R.K.W.), the Intramural Program of the National Human Genome Research Institute (E.A.O.), the Italian Ministry of Environment and the Italian Institute for Environmental Protection and Research (E.R. and C.G.).

REFERENCES

Allendorf FW, Hohenlohe PA, Luikart G (2010). Genomics and the future of conservation genetics. *Nat Rev Genet* **11**: 697-709.

Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**: 1655–1664.

Altobello G (1921). [Fauna of Abruzzo and Molise]. *Mammiferi* **4**: 38–45. [In Italian]

Alvarez RH, Kantarjian HM, Cortes JE (2006). Biology of platelet-derived growth factor and its involvement in disease. *Mayo Clin Proc* **81**: 1241–1257.

Axelsson E, Ratnakuma A, Arendt ML Maqbool K, Webster MT, Perloski M *et al* (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**: 360–364.

Barja I (2009). Prey and prey-age preference by the Iberian wolf *Canis lupus signatus* in a multiple-prey ecosystem. *Wildlife Biol* **15**: 147-154.

Bassem MD, Lars S (2011). Regulation of Human Adipose-Derived Stromal Cell Osteogenic Differentiation by Insulin-Like Growth Factor-1 and Platelet-Derived Growth Factor-alpha. *Plast Reconstr Surg* **127**: 1022-

1023.

Beaumont MA, Balding DJ (2004). Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**: 969–980.

Bocquet-Appel JP (2011). When the world's population took off: the springboard of the Neolithic Demographic Transition. *Science* **333**: 560–561.

Boitani L (2003). Wolf conservation and recovery. In Mech LD & Boitani L, eds. Wolves: Behavior, Ecology, and Conservation. The University of Chicago Press, Chicago: 317-340.

Boyko AR, Quignon P, Li L, Schoenebeck J, Degenhardt JD, Lohmueller KE *et al* (2010). A simple genetic architecture underlies quantitative traits in dogs. *PLoS Biol* **8**: e1000451. doi:10.1371/journal.pbio.1000451.

Cabrera A (1907). Los lobos de Espana. *Bol R Soc Esp Hist Nat* **7**: 193–198.

Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Feled-Alon A, York TL *et al* (2007). Genome-Wide Patterns of Nucleotide Polymorphism in Domesticated Rice. *PLoS Genet* **3**: e163. doi:10.1371/journal.pgen.0030163.

1062 Czarnomska SD, Jędrzejewska B, Borowik T, Niedzialkowska M, Stronen
1063 AV, Nowak S *et al* (2013). Concordant mitochondrial and microsatellite
1064 DNA structuring between Polish lowland and Carpathian Mountain wolves.
1065 *Cons Genet* **14:** 573-588.
1066
1067 Earl DA & Vonholdt BM (2012). Structure Harvester: A website and
1068 program for visualizing structure output and implementing the Evanno
1069 method. *Conserv Genet Resour* **4**: 359-361.
1070
1071 Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of
1072 individuals using the software STRUCTURE: a simulation study. *Mol Ecol*
1073 **14**: 2611-2620.
1074
1075 Fabbri E, Caniglia R, Kusak J, Galov A, Gomerčić T, Haidi Arbanasić H,
1076 Huber D, Randi E (in press). Genetic structure of expanding wolf (*Canis*
1077 *lupus*) populations in Italy and Croatia, and the early steps of the
1078 recolonization of the Eastern Alps. *Mamm Biol.*
1079
1080 Foll M, Gaggiotti O (2008). A Genome-Scan Method to Identify Selected
1081 Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian
1082 Perspective. *Genetics* **180**: 977–993.
1083
1084 Freedman AH, Schweizer RM, Gronau I, Han E, Ortega – Del Vecchyo D,
1085 Silva PM, *et al* (2014). Genome Sequencing Highlights the Dynamic Early

History of Dogs. PLoS Genet. (in press)

Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, and Di Rienzo A (2001). Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am J Hum Genet* **69**: 831–843.

Gattepaille LM, Jakobsson M, Blum MG (2013). Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity* **110**: 409-419.

Gautier M, Vitalis R (2013). Inferring Population Histories Using Genome-Wide Allele Frequency Data. *Mol Biol Evol* **30**: 654–668.

Germonpré M, Sablin MV, Stevens RE, Hedges REM, Hofreiter M, Stiller M, Després VR (2009). Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. *J Arch Science* **36**: 473-490.

Gomercic T, Sindicic M, Galov A, Arbanasic H, Kusak J, Kocijan I *et al*. (2010). High genetic variability of the grey wolf (*Canis lupus L.*) population from Croatia as revealed by mitochondrial DNA control region sequences. *Zool Stud* **49:** 816-823.

1110 Gompper ME (2002). Top carnivores in the suburbs? Ecological and
1111 conservation issues raised by colonization of north eastern North America
1112 by coyotes. *Bioscience* **52**: 185–190.
1113
1114 Gray MM, Sutter NB, Ostrander EA, Wayne RK (2010). The IGF1 small
1115 dog haplotype is derived from Middle Eastern grey wolves. *BMC Biology* **8**:
1116 1-13.
1117
1118 Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003). Novel
1119 multilocus measure of linkage disequilibrium to estimate past effective
1120 population size. *Genome Res* **13**: 635–643.
1121
1122 Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA
1123 (2010). Population Genomics of Parallel Adaptation in Threespine
1124 Stickleback using Sequenced RAD Tags. *PLoS Genet* **6**: e1000862.
1125
1126 Hofreiter M and Barnes I (2010). Diversity lost: are all Holarctic large
1127 mammal species just relict populations? *BMC Biology*, **8**: 46.
1128
1129 Jędrzejewska B, Jędrzejewski W, Bunevich AN, Minkowski L, Okarma H
1130 (1996). Population dynamics of wolves *Canis lupus* in Bialowieza primeval
1131 forest (Poland and Belarus) in relation to hunting by humans, 1847-1993.
1132 *Mammal Rev* **26**: 103-126.
1133

1134 Jędrzejewski W, Branicki W, Veit C, Međugorac I, Pilot M, Bunevich AN
1135 *et al* (2005). Genetic diversity and relatedness within packs in an intensely
1136 hunted population of wolves *Canis lupus*. *Acta Theriol* **50**: 3–22.
1137
1138 Jędrzejewski W, Jędrzejewska B, Andersone-Lilley Z, Balciauskas L,
1139 Mannil P, Ozolins J *et al* (2010). Synthesizing wolf ecology and
1140 management in Eastern Europe: Similarities and contrasts with North
1141 America. In: Musiani M, Boitani L, Paquet PC, editors. The World of
1142 Wolves. New perspectives on ecology, behavior and management. Calgary:
1143 University of Calgary Press. 207–233.
1144
1145 Jones FC, Chan YF, Schmutz J, Grimwood J, Brady S, Southwick A *et al*
1146 (2012) A Genome-wide SNP Genotyping Array Reveals Patterns of Global
1147 and Repeated Species-Pair Divergence in Sticklebacks. *Curr Biol* **22:** 83–
1148 90.
1149
1150 Kalinowski ST (2011) The computer program STRUCTURE does not
1151 reliably identify the main genetic clusters within species: simulations and
1152 implications for human population structure. *Heredity* **106**: 625–632.
1153
1154 Keigwin LD, Donnelly JP, Cook MS, Neal W. Driscoll NW, Brigham-
1155 Grette J (2006). Rapid sea-level rise and Holocene climate in the Chukchi
1156 Sea. *Geology* **34**: 861–864.
1157

1158 Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A *et al*
1159 (2013) Population genomic signatures of divergent adaptation, gene flow
1160 and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes.
1161 Mol Ecol **22:** 2848–2863.
1162
1163 Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M
1164 *et al* (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals
1165 High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol* **10**:
1166 e1001258.
1167
1168 Leonard JA, Vilà C, Fox-Dobbs K, Koch PL, Wayne RK, Van Valkenburgh
1169 B (2007). Megafaunal extinctions and the disappearance of a specialized
1170 wolf ecomorph. *Curr Biol* **17**: 1146-1150.
1171
1172 Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal
1173 M *et al* (2005). Genome sequence, comparative analysis and haplotype
1174 structure of the domestic dog. *Nature* **438**: 803–819.
1175
1176 Lucchini V, Galov A, Randi E (2004). Evidence of genetic distinction and
1177 long-term population decline in wolves (*Canis lupus*) in the Italian
1178 Apennines. *Mol Ecol* **13**: 523-536.
1179
1180 MacNulty DR, Smith DW, Mech LD, Eberly LE (2009). Body size and
1181 predatory performance in wolves: is bigger better? *J Anim Ecol* **78**: 532-

539.

Mattioli L, Capitani C, Gazzola A, Scandura M, Apollonio M (2011). Prey selection and dietary response by wolves in a high-density multi-species ungulate community. *Eur J Wildlife Res* **57**: 909-922.

Mech LD, Seal US (1987). Premature reproductive activity in wild wolves. *J Mammal* **68**: 871-873.

Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F *et al* (2012). Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *PNAS* **109**: E2382–E2390.

Musiani M, Boitani L, Paquet P (2010). The World of Wolves: new perspectives on ecology, behaviour and management. University of Calgary Press, Calgary.

Nowak RM (2003). Wolf evolution and taxonomy. In Mech LD & Boitani L, eds. Wolves: Behavior, Ecology, and Conservation. The University of Chicago Press, Chicago: 239-258.

Nowak RM, Federoff NE (2002). The systematic status of the Italian wolf *Canis lupus. Acta Theriol* **47**: 333–338.

1206 Oliver MK, Piertney SB (2012). Selection Maintains MHC Diversity
1207 through a Natural Population Bottleneck. *Mol Biol Evol* **29**:1713–1720.
1208
1209 Ozolins J, Andersone Z (2001) Status of large carnivore conservation in
1210 the Baltic States. Action plan for the conservation of wolf (*Canis lupus*) in
1211 Latvia. European Commission, Strasbourg, T-PVS (2001) 73 add. 2, Pp. 1-
1212 32.
1213
1214 Patterson N, Price AL, Reich D (2006) Population structure and
1215 eigenanalysis. PLoS Genet 2(12): e190.
1216
1217 Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Bëer E, Robinson S *et al*
1218 (2012). Reliability of genetic bottleneck tests for detecting recent population
1219 declines. *Mol Ecol* **21**: 3403–3418.
1220
1221 Pilot M, Branicki W, Jędrzejewski W, Goszczyński J, Jędrzejewska B,
1222 Dykyy I *et al* (2010). Phylogeographic history of grey wolves in Europe.
1223 *BMC Evol Biol* **21**: 10-104.
1224
1225 Pilot M, Jędrzejewski W, Branicki W, Sidorovich VE, Jędrzejewska B,
1226 Stachura K, Funk SM (2006). Ecological factors influence population
1227 genetic structure of European grey wolves. *Mol. Ecol*. **15**: 4533-4553.
1228
1229 Pritchard J, Stephens M, Donnelly P (2000). Inference of population

structure using multilocus genotype data. *Genetics* **155**: 945-959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575.

Randi E (2011). Genetics and conservation of wolves *Canis lupus* in Europe. *Mammal Rev* **41**: 99-111.

Reddi AH, Cunningham NS (1990). Bone induction by osteogenin and bone morphogenetic proteins. *Biomaterials* **11**: 33-34.

Sastre N (2011). Genética de la conservación: el lobo gris (*Canis lupus*). PhD thesis, Autonomous University of Barcelona, Spain.

Sastre N, Vila C, Salinas M, Bologov VV, Urios V, Sanchez A *et al* (2011). Signatures of demographic bottlenecks in European wolf populations. *Conserv Genet* **12**: 701–712.

Scheet P, Stephens M (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78**: 629-644.

Schmidt MB, Chen EH, Lynch SE (2006). A review of the effects of

1254 insulin-like growth factor and platelet derived growth factor on in vivo
1255 cartilage healing and repair. *Osteoarthr Cartilage* **14**: 403-412.
1256
1257 Sidorovich VE, Tikhomirova LL, Jędrzejewska B (2003). Wolf *Canis lupus*
1258 numbers, diet and damage to livestock in relation to hunting and ungulate
1259 abundance in northeastern Belarus during 1990-2000. *Wildlife Biol* **9**: 103-
1260 111.
1261
1262 Sommer R, Benecke N (2005). Late-Pleistocene and early Holocene history
1263 of canid fauna of Europe (Canidae). *Mamm Biol* **70**: 227-241.
1264
1265 Spiegelhalter DJ, Best NG, Carlin BP, Linde AVD (2002). Bayesian
1266 measures of model complexity and fit. *J Roy Stat Soc B* **64**: 583–639.
1267
1268 Spiridonov G, Spassov N (1985). Wolf - *Canis lupus* L., 1758. In: Botev,
1269 Peshev (ed) Red Data Book of Bulgaria Bulgarian Academy of Science,
1270 Sofia, pp. 132.
1271
1272 Stefani CM, Machado MA, Sallum EA, Sallum AW, Toledo S, Nociti FH Jr
1273 (2000). Platelet-derived growth factor/insulin-like growth factor-1
1274 combination and bone regeneration around implants placed into extraction
1275 sockets: a histometric study in dogs. *Implant Dent* **9**: 126-131.
1276

Stronen AV, Jędrzejewska B, Pertoldi C, Demontis D, Randi E, Niedzialkowska M et al. (2013). North-south differentiation and a region of high diversity in European wolves (*Canis lupus*). PLoS ONE 8: e76454.

Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L *et al* (2007). A single IGF1 allele is a major determinant of small size in dogs. *Science* **316**: 112–115.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Likelihood, Distance, and Parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME et *al* (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Res* **17**: 520–526.

Vilà C (1993) Aspectos morfologicos y ecológicos del lobo iberico *Canis lupus* L. PhD thesis, Universidad de Barcelona, Spain.

Vilà C, Amorim I. R, Leonard J. A, Posada D, Castroviejo J, Petrucci-Fonseca F *et al* (1999). Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. *Mol Ecol* **8**: 2089-2103.

vonHoldt B, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P *et*

al (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**: 898-902.

vonHoldt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H *et al* (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res*. **21**: 1294-1305.

Zhao S, Pingping Z, Dong S, Zhan X, Wu Q, Guo X *et al* (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet* **45**: 67-71.

Zulliger D, Schnyder E, Gugerli F (2013). Are adaptive loci transferable across genomes of related species? Outlier and environmental association analyses in Alpine Brassicaceae species. *Mol Ecol* **22**:1626-1639.

Table 1. Genetic differentiation among grey wolf populations and coyotes calculated in EIGENSOFT based on (A) the 34K dataset, and (B) the loci putatively under differential selection in European populations. Above the diagonal: average divergence between populations; On the diagonal: average divergence within populations; Below the diagonal: pair-wise F_{ST} between populations. All the pair-wise differences were significant (ANOVA, $P<0.05$).

	Grey wolves					
	Italy	Iberian Peninsula	Eastern Europe	Asia	North America	Coyotes
(A)						
Italy	0.609	1.165	1.155	1.526	1.337	1.539
Iberian Peninsula	*0.293*	0.871	1.177	1.534	1.356	1.564
Eastern Europe	*0.195*	*0.128*	1.098	1.477	1.311	1.513
Asia	*0.229*	*0.167*	*0.059*	1.623	1.584	1.745
North America	*0.284*	*0.221*	*0.114*	*0.112*	1.095	1.505
Coyotes	*0.467*	*0.404*	*0.296*	*0.275*	*0.305*	0.706
(B)						
Italy	1.274	3.509	4.856	4.484	4.364	4.288
Iberian Peninsula	*0.925*	0.923	2.761	2.732	2.490	2.376
Eastern Europe	*0.848*	*0.758*	0.938	1.233	1.136	1.559
Asia	*0.796*	*0.694*	*0.086*	1.290	1.279	1.605
North America	*0.821*	*0.711*	*0.170*	*0.110*	1.017	1.524
Coyotes	*0.867*	*0.751*	*0.422*	*0.359*	*0.444*	0.559

1326 Table 2. Heterozygosity, linkage disequilibrium and autozygosity in local
1327 wolf populations from Europe. The extent of linkage disequilibrium is
1328 measured as the average distance between loci at which r^2 falls below 0.5.
1329 Autozygosity is measured as average number of homozygous segments per
1330 individual, and their average length.

Local population	H_O (SE)	H_E (SE)	Distance (r^2<0.5) (Kb)	Average No. of homozygous segments per individual (SE)	Average length of homozygous segments (Kb) per segment per ind. (SE)
S Poland and S Belarus	0.235 (0.0010)	0.232 (0.0008)	5.00	2.7 (0.6)	3634 (432)
N Poland	0.234 (0.0010)	0.220 (0.0008)	10.00	2.2 (0.3)	2950 (631)
N Belarus	0.233 (0.0010)	0.263 (0.0010)	3.75	5.0 (1.1)	4378 (1105)
NE Russia	0.232 (0.0010)	0.219 (0.0008)	3.75	3.3 (0.8)	3696 (1107)
Kirov Region, Russia	0.230 (0.0010)	0.233 (0.0008)	7.50	2.4 (0.9)	3080 (773)
S Russia and E Ukraine	0.228 (0.0009)	0.233 (0.0008)	2.50	4.7 (1.1)	3502 (844)
Balkans	0.217 (0.0010)	0.223 (0.0008)	10.00	3.2 (1.5)	1771 (631)
Carpathians	0.214 (0.0010)	0.257 (0.0010)	7.50	6.5 (0.7)	5142 (399)
Iberian Peninsula	0.173 (0.0010)	0.169 (0.0008)	275.00	1.5 (0.8)	1902 (928)
Italy	0.161 (0.0010)	0.155 (0.0010)	>1000.00	1.6 (0.7)	2449 (937)

Figure legends

Figure 1. (A) Map of sample distribution. The red circles represent sampling locations, which are exact, except for the Iberian Peninsula, where the exact sample locations were unknown. A sampling location may be shared by several individuals. The range of the grey wolf is marked in pink on the main map, and in red on the small map showing the worldwide distribution of this species; (B) Subdivision into local populations in Eastern Europe based on the geographical proximity of the samples and data on their genetic differentiation from Pilot et al. (2006). Local populations with sample size at least 5 individuals were used in the LD decay and ROH analyses.

Figure 2. Results of (A) ADMIXTURE and (B) STRUCTURE clustering analysis of European wolf populations in comparison with other wolf populations and the coyotes, for K=6 and K=7. The analysis was performed for the LD-pruned 34K SNP set. Within Eastern Europe, the samples are sorted according to their geographical locations, from the Kirov Region in Russia on the left to the Balkans and Carpathians on the right.

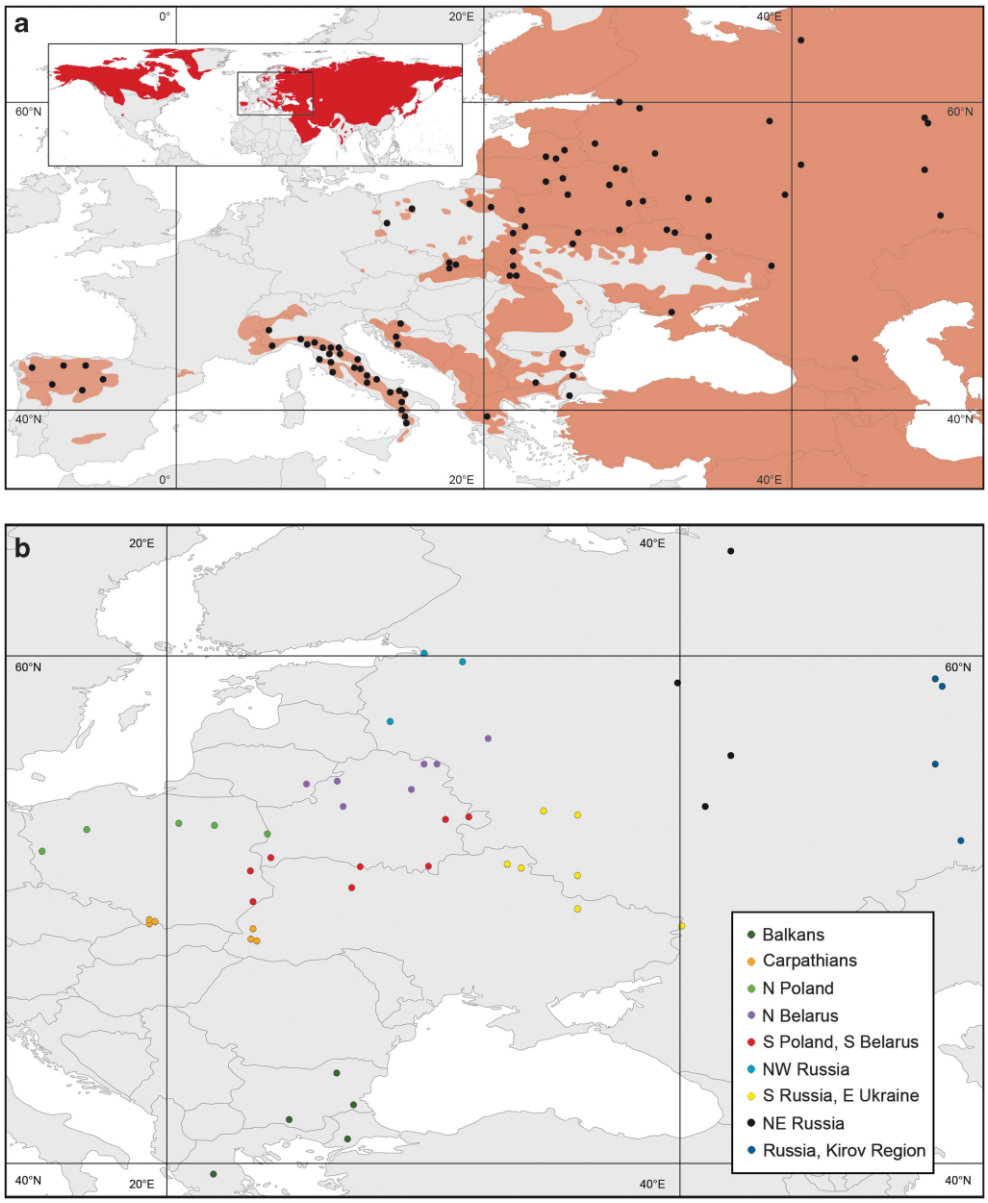
Figure 3. Principal component analysis illustrating the extent of genetic diversification at the genome-wide 34K SNP set (A-C), and at loci putatively under diversifying selection in European wolves (D, E) among the following populations: (A) and (D) European, Asian and North American grey wolves, and coyotes; (B) and (E) European grey wolves; and C) Eastern European grey wolves.

Figure 4. Evolutionary relationships among X chromosome haplotypes of females inferred using the neighbour-joining method. The distances were computed using the p-distance measure. Bootstrap support is shown if higher than 50% of 1000 replicates.

Figure 5. (A) Extent of linkage disequilibrium in European wolf populations. Average genotypic association coefficient r^2 is presented as a function of inter-SNP distance for each local wolf population. (B) Frequency distribution of runs of homozygosity in European wolf populations. (C) Temporal changes of N_E in European wolves, with North American wolves presented for a comparison.

Figure 6. Signatures of selection in the Iberian, Italian and Eastern European wolf populations inferred using the program BAYESCAN. The vertical axis indicates mean F_{ST} values between each of the three populations, and the horizontal axis indicates the logarithm of posterior odds (log(PO)). The vertical line indicates the log(PO) value corresponding to the false discovery rate threshold of 0.05. Loci on the right of this line are putatively under selection.

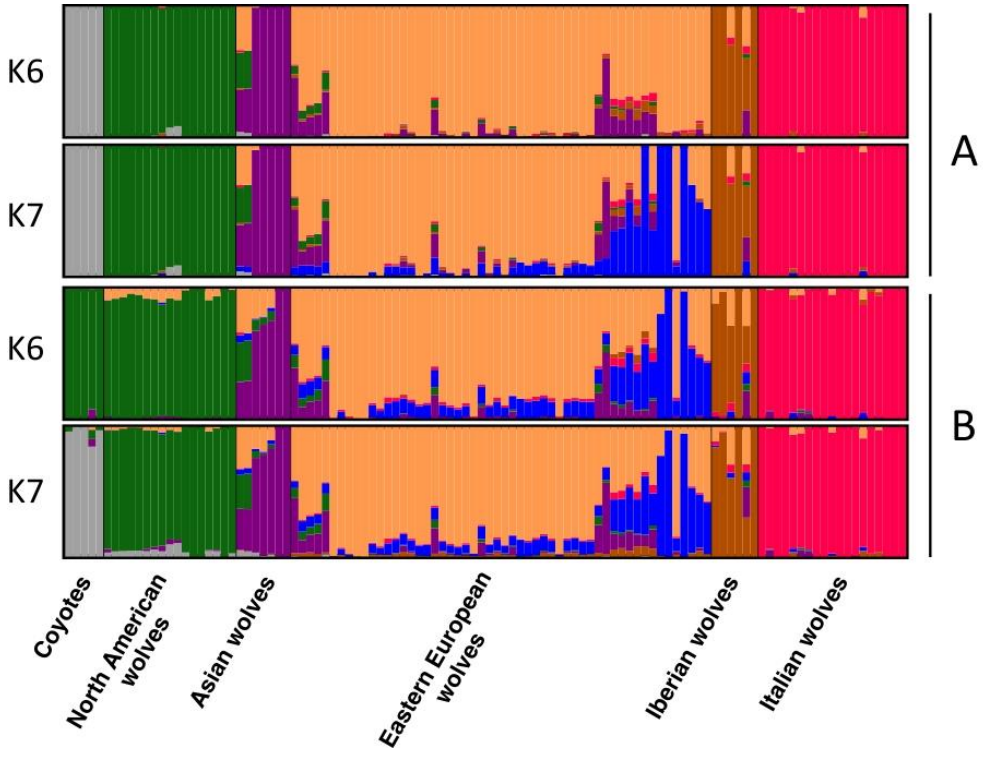
1377 Figure 1



1378

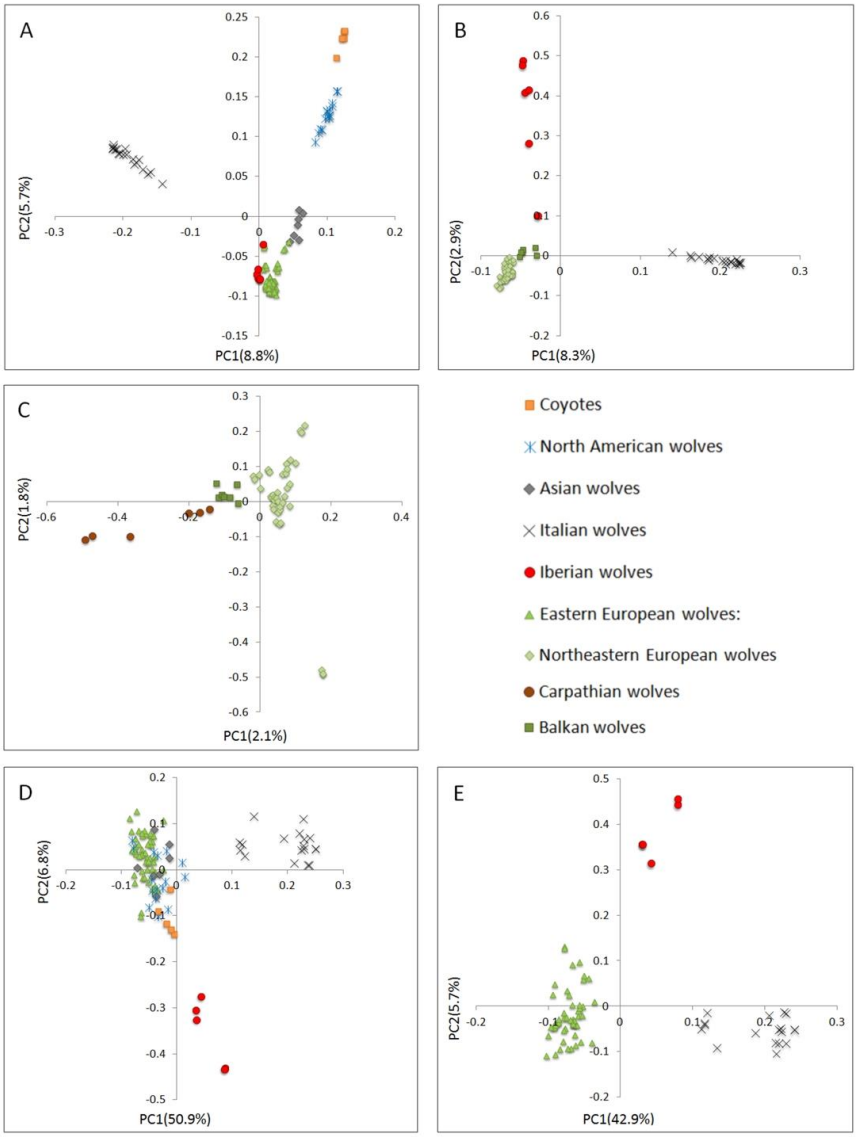
1379

1380 Figure 2



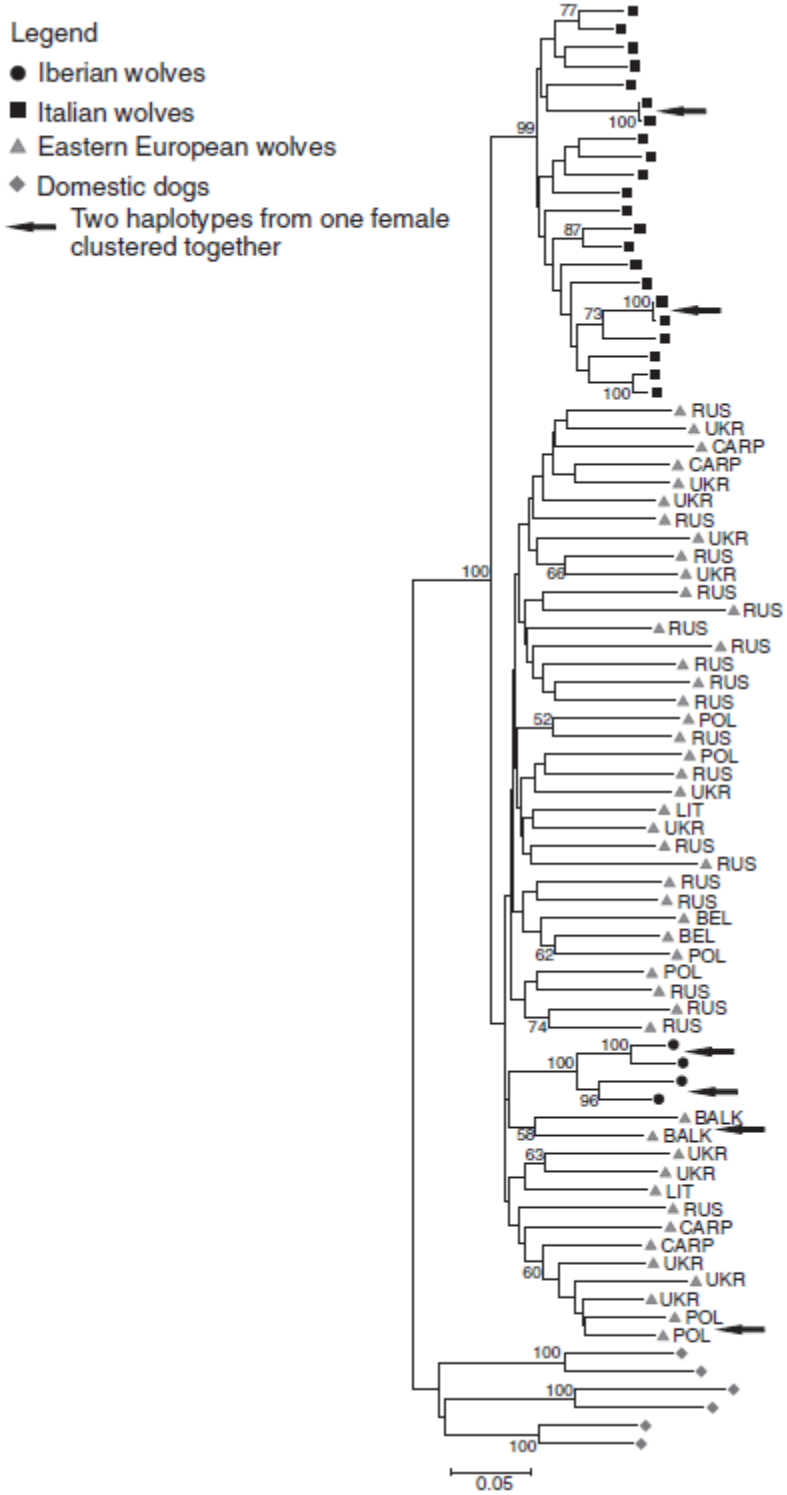
1381

1382 Figure 3



1383

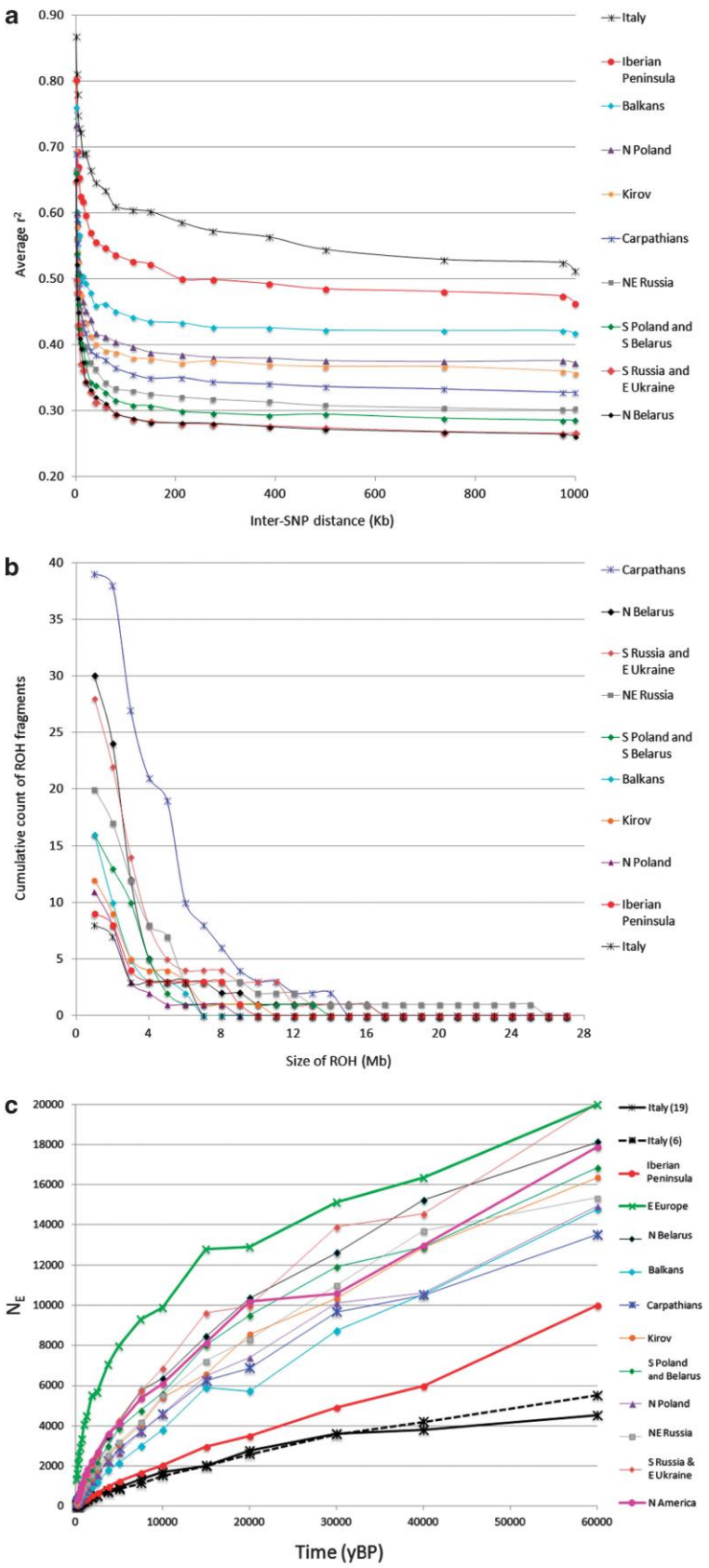
1384 Figure 4



1385

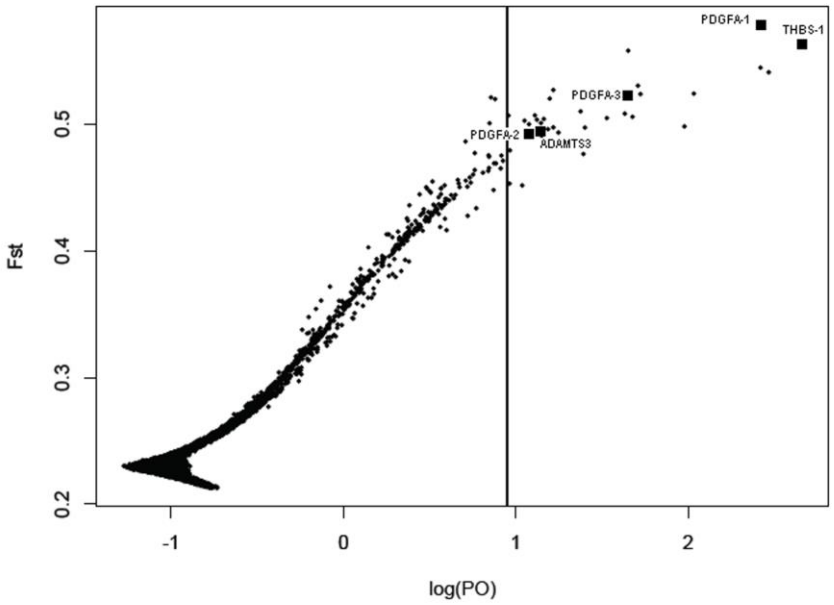
1386

1387 Figure 5



1388

1389 Figure 6



1390